# EVALUATION MEASURE FOR CLASSIFICATION

- True Positive, True Negative, False Positive, False Negative

| | | **Predicted class** | |
|---|---|---|---|
| | | Yes | No |
| **Actual class** | Yes | TP: True positive | FN: False negative |
| | No | FP: False positive | TN: True negative |

- Machine Learning methods usually minimize FP+FN
- In practice FP and FN could have different costs
  - Medical diagnostic tests: does X have leukemia?
  - Loan decisions: approve mortgage for X?

# EVALUATION MEASURE FOR CLASSIFICATION

- Multi-class problems:

| predicted → / real ↓ | *Class_1* | *Class_2* | *Class_3* |
|---|---|---|---|
| *Class_1* | 94 | 16 | 10 |
| *Class_2* | 21 | 113 | 16 |
| *Class_3* | 4 | 4 | 92 |

# EVALUATION MEASURE FOR CLASSIFICATION

- Accuracy $\frac{TP + TN}{TP + TN + FP + FN}$
- Precision $\frac{TP}{TP + FP}$
- Recall $\frac{TP}{TP + FN}$
- *F-measure* $\frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$

Traditional (Global) Performance Measures

# EVALUATION MEASURE FOR CLASSIFICATION

Given a label *l* belonging to the set of labels L:

- Precision $$P(l) = \frac{\#\text{ of instances correctly predicted as l}}{\#\text{ of instances predicted as l}}$$
- Recall $$R(l) = \frac{\#\text{ of instances correctly predicted as l}}{\#\text{ of instances of class l}}$$
- *F-measure* $$F(l) = \frac{2 \cdot P(l) \cdot R(l)}{P(l) + R(l)}$$

Class Level Performance Measures

# EVALUATION MEASURE FOR CLASSIFICATION

Given a label *l* belonging to the set of labels L:

Macro-average

$$Perf^{*} = \frac{1}{|L|}\sum_{l=1}^{|L|} Perf(l)$$

All classes are equally important

Micro-average

$$Perf^{*} = \sum_{l=1}^{|L|} \frac{|class(l)|}{|\#\ of\ instances|} Perf(l)$$

Predominant classes are more important

# EVALUATION MEASURE FOR CLASSIFICATION

- ROC curve: Receiver Operating Characteristic
  - graphical plot that shows the performance of a classifier as its discrimination threshold is varied
  - true positive rate vs false positive rate (at various **threshold** settings)